Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02021650.3

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Europäisches Patentamt | European Patent Office | Office européen des brevets

Anmeldung Nr:
Application no.: 02021650.3
Demande no:

Anmeldetag:
Date of filing: 27.09.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Roche Vitamins AG

4070 Basel
SUISSE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

A gene encoding aldehyde dehydrogenase and use thereof

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N9/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

Roche Vitamins AG, CH-4070 Basle, Switzerland

Case 21424

A gene encoding aldehyde dehydrogenase and use thereof

The present invention relates to a novel DNA which encodes aldehyde dehydrogenase (SNDH) derived from *Gluconobacter oxydans* DSM 4025, an expression vector containing the DNA and recombinant organisms containing the expression vector. Furthermore, the present invention concerns a process for producing recombinant aldehyde dehydrogenase protein and a process for producing L-ascorbic acid (vitamin C) and/or 2-keto-L-gulonic acid (2-KGA) from L-sorbosone by using the recombinant aldehyde dehydrogenase protein or recombinant organisms containing the expression vector.

Vitamin C is one of indispensable nutrient factors for human beings and has been commercially synthesized by the Reichstein process for about 60 years. Synthetic vitamin C is also used in animal feeds even though farm animals can synthesize it in their own body. Although the Reichstein process has many advantageous points for industrial vitamin C production, it still has undesirable problems such as high energy consuming and usage of considerable quantities of organic and inorganic solvents. Therefore, over the past decades, many approaches to manufacture vitamin C using enzymatic conversions, which would be more economical as well as ecological, have been investigated.

The present invention provides a gene coding for an aldehyde dehydrogenase (SNDH), e.g. from *G. oxydans* DSM 4025 (FERM BP-3812) as disclosed in US 6,242,233, catalyzing the conversion of L-sorbosone not only to 2-keto-L-gulonic acid (2-KGA), but also to vitamin C.

The present invention provides a novel DNA which encodes aldehyde dehydrogenase (SNDH) derived from *G. oxydans* DSM 4025. The present invention also provides an expression vector containing the DNA and recombinant organisms containing the expression vector. Furthermore, the present invention provides a process for producing recombinant aldehyde dehydrogenase protein and a process for producing L-ascorbic acid

Hei/fm, 27.09.2002

(vitamin C) and/or 2-keto-L-gulonic acid (2-KGA) from L-sorbosone by using the recombinant aldehyde dehydrogenase protein or recombinant organisms containing the expression vector.

This invention is directed to a nucleic acid molecule comprising a nucleotide sequence which encodes a protein having the amino acid sequence at positions 32-578 of SEQ ID NO:2, or a protein derived from that protein by substitution, deletion, insertion or addition of one or more amino acids in the amino acid sequence at positions 32-578 of SEQ ID NO:2, which has the SNDH activity. This invention is also directed to an expression vector comprising such a polynucleotide, especially one which functions in a host cell belonging to bacterial cells, yeast cells and plant cells. Preferably the host cell belongs to the genera *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. This invention is also directed to a recombinant organism comprising such an expression vector, especially one which has the polynucleotide on its chromosomal DNA, and preferably one which is a microorganism belonging to the genera *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*.

This invention is also directed to a nucleic acid molecule as described above consisting of a polynucleotide comprising the nucleotide sequence at positions 351-2084 of SEQ ID NO:1, or a polynucleotide comprising the nucleotide at positions 258-2084 of SEQ ID NO:1, or a polynucleotide capable of hybridizing to the above polynucleotides, and which encodes a protein having SNDH activity.

This invention includes a nucleic acid molecule which comprises the nucleotide sequence at positions 351-2084 of SEQ ID NO:1. This invention is also directed to an expression vector comprising such a polynucleotide, especially one which functions in a host cell belonging to bacterial cells, yeast cells and plant cells. Preferably the host cell belongs to the genera *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. This invention is also directed to a recombinant organism comprising such an expression vector, especially one which has the polynucleotide on its chromosomal DNA, and preferably one which is a microorganism belonging to the genera *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*.

This invention is also directed to a process for producing 2-KGA more efficiently by using a disruptant of the SNDH gene of *G. oxydans* DSM 4025 (FERM BP-3812).

As used herein, "cloning vector" means a plasmid or phage DNA or other DNA sequence which is able to replicate autonomously in a host cell, and which is characterized by one or

a small number of restriction endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a DNA fragment may be spliced in order to bring about its replication and cloning. The cloning vector may further contain a marker suitable for use in the identification of cells transformed with the cloning vector. Markers, e.g., provide tetracycline resistance or ampicillin resistance.

As used herein, "expression" refers to the process by which a polypeptide is produced from a structural gene. The process involves transcription of the gene into mRNA and the translation of such mRNA into polypeptide(s).

As used herein, "expression vector" means a vector similar to a cloning vector but which is capable of enhancing the expression of a gene that has been cloned into it, after transformation into a host. The cloned gene is usually placed under the control of (i.e., operably linked to) certain control sequences such as promoter sequences. Promoter sequences may be either constitutive or inducible.

As used herein, "gene" refers to a DNA sequence that contains information needed for expressing a polypeptide or protein.

As used herein, "nucleic acid molecule" includes both DNA and RNA and, unless otherwise specified, includes both double-stranded, single-stranded nucleic acid, and nucleosides thereof. Also included are hybrids such as DNA-RNA hybrids, DNA-RNA-protein hybrids, RNA-protein hybrids, and DNA-protein hybrids.

As used herein, "host" includes any prokaryotic or eukaryotic cell that is the recipient of a replicable expression vector or cloning vector. A "host," as the term is used herein, also includes prokaryotic or eukaryotic cells that can be genetically engineered by well known techniques to contain desired gene(s) on its chromosome or genome. Examples of such hosts are known to the skilled artisan.

As used herein, "mutation" refers to a single base pair change, insertion or deletion in the nucleotide sequence of interest.

As used herein, "mutagenesis" refers to a process whereby a mutation is generated in DNA. With "random" mutagenesis, the exact site of mutation is not predictable, occurring anywhere in the chromosome of the microorganism, and the mutation is brought about as a result of physical damage caused by agents such as radiation or chemical treatment.

As used herein, "operon" refers to a unit of bacterial gene expression and regulation, including the structural genes and regulatory elements in DNA.

As used herein, "phenotype" refers to observable physical characteristics dependent upon the genetic constitution of a microorganism.

As used herein, "promoter" means DNA sequence generally described as the 5' region of a gene, located proximal to the start codon. The transcription of an adjacent gene(s) is initiated at the promoter region. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter.

As used herein, "recombinant" means a recombinant host which may be any prokaryotic or eukaryotic cell and contains the desired cloned gene(s) on an expression vector or cloning vector. This term also include those prokaryotic or eukaryotic cells that have been genetically engineered to contain the desired gene(s) in the chromosome or genome of that organism.

As used herein, " recombinant vector" includes any cloning vector or expression vector which contains the desired cloned gene(s).

As used herein, "SNDH" stands for L-sorbosone dehydrogenase and "RHA" stands for RNA helicase A.

As used herein, "percent identical" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical amino acids in the compared amino acid sequence by a sequence analysis program as exemplified below.

The invention provides an isolated nucleic acid molecule encoding the enzyme (SNDH). Methods and techniques designed for the manipulation of isolated nucleic acid molecules are well known in the art. Methods for the isolation, purification, and cloning of nucleic acid molecules, as well as methods and techniques describing the use of eukaryotic and prokaryotic host cells and nucleic acid and protein expression therein, are known to the skilled person.

Functional derivatives are defined on the basis of the amino acid sequences of the present invention by addition, insertion, deletion and/or substitution of one or more amino acid residues of such sequences wherein such derivatives still have the SNDH activity measured by an assay known in the art or specifically described herein. Such functional derivatives can be made either by chemical peptide synthesis known in the art or by recombinant

techniques on the basis of the DNA sequences as disclosed herein by methods known in the state of the art. Amino acid exchanges in proteins and peptides which do not generally alter the activity of such molecules are known in the state of the art.

In particular embodiments of the present invention, conservative substitutions of interest occur as follows: As example substitutions, Ala to Val/Leu/Ile, Arg to Lys/Gln/Asn, Asn to Gln/His/Lys/Arg, Asp to Glu, Cys to Ser, Gln to Asn, Glu to Asp, Gly to Pro/Ala, His to Asn/Gln/Lys/Arg, Ile to Leu/Val/Met/Ala/Phe/norLeu, Lys to Arg/Gln/Asn, Met to Leu/Phe/Ile, Phe to Leu/Val/Ile/Ala/Tyr, Pro to Ala, Ser to Thr, Thr to Ser, Trp to Tyr/Phe, Tyr to Trp/Phe/Thr/Ser, and Val to Ile/Leu/Met/Phe/Ala/norLeu are reasonable. As preferred examples, Ala to Val, Arg to Lys, Asn to Gln, Asp to Glu, Cys to Ser, Gln to Asn, Glu t Asp, Gly to Ala, His to Arg, Ile to Leu, Leu to Ile, Lys to Arg, Met to Leu, Phe to Leu, Pro to Ala, Ser to Thr, Thr to Ser, Trp to Tyr, Tyr to Phe, and Val to Leu are reasonable. If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions described above, are introduced and the products screened.

Unless otherwise mentioned, all amino acid sequences determined by sequencing the purified SNDH protein herein were determined using an automated amino acid sequencer (such as model 470A, Perkin-Elmer Applied Biosystems).

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the model ALF express II, Amersham Pharmacia Biotech), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of the DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

Furthermore the present invention is directed to polynucleotides encoding polypeptide having the SNDH activity as disclosed in the sequence listing as SEQ ID NO:2 as well as the complementary strands, or those which include these sequences, DNA sequences or fragments thereof, and DNA sequences, which hybridize under standard conditions with such sequences but which encode for polypeptides having exactly the same amino acid sequence.

Standard conditions for hybridization mean in this context the conditions which are generally used by a person skilled in the art to detect specific hybridization signals, or preferably so called "stringent hybridization conditions" used by a person skilled in the art.

Thus, as used herein, the term "stringent hybridization conditions" means hybridization will occur if there is 95% and preferably at least 97% identity between the sequences. Stringent hybridization conditions are, e.g., conditions under over night incubation at 42°C using a digoxigenin (DIG)-labeled DNA probe (constructed by using a DIG labeling system; Roche Dignostics GmbH, 68298 Mannheim, Germany) in a solution comprising 50% formamide, 5 × SSC (150 mM NaCl, 15 mM trisodium citrate), 0.2% sodium dodecyl sulfate, 0.1% N-lauroylsarcosine, and 2% blocking reagent (Roche Dignostics GmbH), followed by washing the filters in 0.1× SSC at about 60°C.

Briefly, the SNDH gene, the DNA molecule containing said gene, the recombinant expression vector and the recombinant organism used in the present invention can be obtained by the following steps:
(1) Isolating chromosomal DNA from *G. oxydans* DSM 4025 and constructing the gene library with the chromosomal DNA in an appropriate host cell, e.g. *E. coli*.
(2) Cloning the SNDH gene from a chromosomal DNA by colony-, plaque-, or Southern-hybridization, PCR (polymerase chain reaction) cloning, western-blot analysis and so on.
(3) Determining the nucleotide sequence of the SNDH gene obtained as above by conventional methods to select DNA molecule containing said SNDH gene and constructing the recombinant expression vector on which SNDH gene can express efficiently.
(4) Constructing recombinant organisms carrying SNDH gene by an appropriate method for introducing DNA into host cell, e.g. transformation, transduction, conjugal transfer and/or electroporation, which host cell thereby becomes a recombinant organism of this invention.

The materials and the techniques used in the above aspect of the present invention are exemplified in detail as follows:

A total chromosomal DNA can be purified by a procedure well known in the art. The desired gene can be cloned in either plasmid or phage vectors from a total chromosomal DNA typically by either of the following illustrative methods:

(i) The partial amino acid sequences are determined from the purified proteins or peptide fragments thereof. Such whole protein or peptide fragments can be prepared by the isolation of such a whole protein or by peptidase-treatment from the gel after SDS-polyacrylamide gel electrophoresis. Thus obtained protein or fragments thereof are applied to protein sequencer such as Applied Biosystems automatic gas-phase sequencer 470A. The amino acid sequences can be utilized to design and prepare oligonucleotide probes and/or primers with DNA synthesizer such as Applied Biosystems automatic DNA sequencer 381A. The probes can be used for isolating clones carrying the target gene from a gene library of the strain carrying the target gene by means of Southern-, colony- or plaque-hybridization.

(ii) Alternatively, for the purpose of selecting clones expressing target protein from the gene library, immunological methods with antibody prepared against the target protein can be applied.

(iii) The DNA fragment of the target gene can be amplified from the total chromosomal DNA by PCR method with a set of primers, i.e. two oligonucleotides synthesized according to the amino acid sequences determined as above. Then a clone carrying the target-whole gene can be isolated from the gene library constructed, e.g. in *E. coli* by Southern-, colony-, or plaque-hybridization with the PCR product obtained above as the probe.

DNA sequences which can be made by PCR by using primers designed on the basis of the DNA sequences disclosed therein by methods known in the art are also an object of the present invention.

Above mentioned antibody can be prepared with the purified SNDH proteins, the purified recombinant SNDH proteins such as His-tagged SNDH expressed in *E. coli*, or its peptide fragment as an antigen.

Once a clone carrying the desired gene is obtained, the nucleotide sequence of the target gene can be determined by a well known method such as dideoxy chain termination method with M13 phage.

The gene encoding the L-sorbosone dehydrogenase activity of the present invention is illustrated in Figure 1, showing a restriction map of the SNDH and ORF-A genes wherein ORF means open reading frame and Signal seq. means the putative signal peptide sequence of the SNDH gene, and Figure 2, showing a physical map of the insert DNA fragments of

cosmid pVSN5, and pUC plasmids pUCSNP4, pUCSNP9, pUCSN19, and pUCSN5. In the physical map of pVSN5, the arrow filled in gray shows the SNDH gene.

This specific gene encodes the SNDH enzyme having 578 amino acid residues together with a putative signal peptide of 31 amino acid residues (SEQ ID NO:2). In terms of nucleotide sequences, the coding region of the SNDH gene is positions at 258-2087 of SEQ ID NO:1 and includes the signal peptide (258-350) and the stop codon (2085-2087). Thus the sequence without the stop codon is positions at 258-2084 of SEQ ID NO:1, and additionally without the signal sequence is positions at 351-2084 of SEQ ID NO:1.

To express the desired gene/nucleotide sequence isolated from *G. oxydans* DSM 4025 efficiently, various promoters can be used; e.g., the original promoter of the gene, promoters of antibiotic resistance genes such as kanamycin resistant gene of Tn5, ampicillin resistant gene of pBR322, and beta-galactosidase of *E. coli* (lac), trp-, tac-, trc-promoter, promoters of lambda phage and any promoters which can be functional in a host cell. For this purpose, the host cell can be selected from a group consisting of bacterial cells, plant cells, and yeast cells. Preferably the host cell belongs to the genera *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. Among the preferred host cells, most preferably it belongs to *G. oxydans*, e.g., *G. oxydans* DSM 4025 (FERM BP-3812), which had been deposited as DSM 4025 on Mar. 17, 1987 under the conditions of the Budapest Treaty at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany.

For expression, other regulatory elements, such as a Shine-Dalgarno (SD) sequence (e.g., AGGAGG and so on including natural and synthetic sequences operable in the host cell) and a transcriptional terminator (inverted repeat structure including any natural and synthetic sequence operable in the host cell) which are operable in the host cell (into which the coding sequence will be introduced to provide a recombinant cell of this invention) can be used with the above described promoters.

For the expression of polypeptides which locate in periplasmic space, like the SNDH protein of the present invention, a signal peptide, which contains usually 15 to 50 amino acid residues and is totally hydrophobic, is preferably associated. A DNA encoding a signal peptide can be selected from any natural and synthetic sequence operable in the desired host cell.

A wide variety of host/cloning vector combinations may be employed in cloning the double stranded DNA. Preferred vectors for the expression of the gene of the present

invention in *E. coli* is selected from any vectors usually used in *E. coli*, such as pQE vectors which can express His-tagged recombinant proteins (QIAGEN K.K., Tokyo, Japan), pBR322 or its derivatives including pUC18 and pBluescript II (Stratagene Cloning Systems, Calif., USA), pACYC177 and pACYC184 and their derivatives, and a vector derived from a broad host range plasmid such as RK2 and RSF1010. A preferred vector for the expression of the nucleotide sequence of the present invention in bacteria including *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter* or *Klebsiella* is selected from any vectors which can replicate in *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, and/or *Klebsiella* as well as in a preferred cloning organism such as *E. coli*. The preferred vector is a broad-host-range vector such as a cosmid vector like pVK100 and its derivatives and RSF1010. Copy number and stability of the vector should be carefully considered for stable and efficient expression of the cloned gene and also for efficient cultivation of the host cell carrying the cloned gene. Nucleic acid molecules containing transposable elements such as Tn5 can also be used as a vector to introduce the desired gene into the preferred host, especially on a chromosome. Nucleic acid molecules containing any DNAs isolated from the preferred host together with the gene of the present invention are also useful to introduce this gene into the preferred host cell, especially on a chromosome. Such nucleic acid molecules can be transferred to the preferred host by applying any of a conventional method, e.g., transformation, transduction, conjugal mating or electroporation, which are well known in the art, considering the nature of the host cell and the nucleic acid molecule.

The SNDH gene/nucleotide sequences provided in this invention are ligated into a suitable vector containing a regulatory region such as a promoter, a ribosomal binding site, and a transcriptional terminator operable in the host cell described above with a well-known method in the art to produce an expression vector.

To construct a recombinant microorganism carrying a recombinant expression vector, various gene transfer methods including transformation, transduction, conjugal mating, and electroporation can be used. The method for constructing a recombinant cell may be selected from the methods well-known in the field of molecular biology. Conventional transformation systems can be used for *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. A transduction system can also be used for *E. coli*. Conjugal mating system can be widely used in Gram-positive and Gram-negative bacteria including *E. coli*, *P. putida*, and *Gluconobacter*. An example of conjugal mating is disclosed in WO 89/06,688. The conjugation can occur in liquid medium or on a solid surface. Examples for a recipient for SNDH production include microorganisms of *Gluconobacter*,

Acetobacter, Pseudomonas, Acinetobacter, Klebsiella or *Escherichia*. To the recipient for conjugal mating, a selective marker may be added; e.g., resistance against nalidixic acid or rifampicin is usually selected. Natural resistance can also be used; e.g., resistance against polymyxin B is useful for many *Gluconobacters*.

The present invention provides recombinant SNDH. One can increase the production yield of the SNDH enzyme by introducing the SNDH gene provided by the present invention into a host cell including *G. oxydans* DSM 4025. One can also produce more efficiently the SNDH proteins in a host cell selected from a group consisting of *Gluconobacter, Acetobacter, Pseudomonas, Acinetobacter, Klebsiella* or *Escherichia* by using the SNDH gene of the present invention. The microorganism may be cultured in an aqueous medium supplemented with appropriate nutrients under aerobic conditions. The cultivation may be conducted at a pH of 4.0 to 9.0, preferably 6.0 to 8.0. The cultivation period varies depending on the pH, temperature and nutrient medium to be used, and is preferably about 1 to 5 days. The preferred temperature range for carrying out the cultivation is from about 13°C to about 36°C, preferably from about 18°C to about 33°C. It is usually required that the culture medium contains such nutrients as assimilable carbon sources, e.g., glycerol, D-mannitol, D-sorbitol, erythritol, ribitol, xylitol, arabitol, inositol, dulcitol, D-ribose, D-fructose, D-glucose, and sucrose, preferably D-sorbitol, D-mannitol, and glycerol; and digestible nitrogen sources such as organic substances, e.g., peptone, yeast extract, baker's yeast, urea, amino acids, and corn steep liquor. Various inorganic substances may also be used as nitrogen sources, e.g., nitrates and ammonium salts. Furthermore, the culture medium usually contains inorganic salts, e.g., magnesium sulfate, potassium phosphate, and calcium carbonate.

An embodiment for the isolation and purification of the recombinant SNDH from the microorganism after the cultivation is briefly described hereinafter: Cells are harvested from the liquid culture broth by centrifugation or filtration. The harvested cells are washed with water, physiological saline of a buffer solution having an appropriate pH. The washed cells are suspended in the buffer solution and disrupted by means of a homogenizer, sonicator or French press, or by treatment with lysozyme and so on to give a solution or disrupted cells. The recombinant SNDH is isolated and purified from the cell-free extract or disrupted cells, preferably from the cytosol fraction of the microorganism.
The recombinant SNDH can be immobilized on a solid carrier for solid phase enzyme reaction. The present invention also provides recombinant cells. The recombinant cell can produce vitamin C and/or 2-KGA from L-sorbosone with the recombinant organisms.

In one embodiment, the invention provides a process for the disruption of the gene by classical mutagenesis by agents such as UV-irradiation or chemical treatment by any mutation reagents, e.g., *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG), ICR170, acrydine orange, and so on, *in vivo* as well as *in vitro*.

In another embodiment, the invention provides a process for the disruption of the gene by DNA recombination techniques such as transposon insertion, site directed mutagenesis by PCR, and so on, *in vivo* as well as *in vitro*.

In another embodiment, the invention provides a process for producing 2-KGA using the disruptants described above by fermentation from appropriate substrate, e.g., L-sorbosone, L-sorbose, and D-sorbitol, in appropriate equipment such as jar fermentors, flasks, and tubes. Also the invention provides a process for producing 2-KGA using cell free extract of the disruptants described above by incubation from appropriate substrate, e.g., L-sorbosone, L-sorbose, and D-sorbitol, in appropriate equipment such as a bioreactor, and etc.

Example 1: Amino acid sequencing from the N-terminus of SNDH

Partial amino acid sequence of the N-terminal 75 kDa subunit of the SNDH protein was determined (SEQ ID NO:3). About 10 μg of the SDS-treated purified SNDH (consisted of 75 kDa subunits) was subjected to SDS-PAGE, and the protein band was electroblotted onto a PVDF membrane. The protein blotted on the membrane was soaked in a digestion buffer (100 mM potassium phosphate buffer, 5 mM dithiothreitol, and 10 mM EDTA, pH 8.0) and incubated with 5.04 μg of pyroglutamate aminopeptidase (SIGMA, USA) at 30°C for 24 hours. After incubation, the membrane was washed with deionized water and subjected to N-terminal amino acid sequencing using an automated amino acid sequencer (ABI model 490, Perkin Elmer Corp., Conn., USA). As a result, 14 residues of the N-terminal amino acid sequence was obtained as illustrated in SEQ ID NO:3.

Example 2: Cloning of partial SNDH gene by PCR

Amplification of partial SNDH gene fragment was carried out by PCR with chromosomal DNA of *G. oxydans* DSM 4025 (FERM BP-3812) and degenerate oligonucleotide DNA primers, P11 (SEQ ID NO:6) and P12 (SEQ ID NO:7). Both of the primers were degenerate DNA mixtures having bias for *Gluconobacter* codon usage. The PCR was performed with thermostable taq polymerase (TAKARA Ex Taq™, Takara Shuzo Co., Ltd., Seta 3-4-1, Otsu, Shiga, 520-2193, Japan), using a thermal cycler (Gene Amp PCR System 2400-R, PE Biosytems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA). The reaction mixture (25 μl) consisted of 200 μM of dNTPs, 50 pmol of each primer (24 ~ 48 degenera-

cy), 5 ng of the chromosomal DNA, and 1.25 units of the DNA polymerase in the buffer provided from the supplier. The reaction was carried out with 5 cycles of 1) denaturation step at 94°C for 30 sec; 2) annealing step at 37°C for 30 sec; 3) synthesis step at 70°C for 1 min plus 25 cycles of 1) denaturation step at 94°C for 30 sec; 2) annealing step at 50°C for 30 sec; 3) synthesis step at 70°C for 1 min. As a result, 41 bp DNA fragment was specifically amplified and cloned into vector pCR 2.1-TOPO (Invitrogen, 1600 Faraday Avenue Carlsbad, California 92008, USA) to obtain a recombinant plasmid pMTSN2. The cloned 41 bp-DNA which encodes N-terminal partial amino acid sequence of the mature SNDH protein, was confirmed the nucleotide sequence by dideoxy-chain termination method (F. Sanger et al, Proc. Natl. Acad. Sci. USA, 74, 5463-5467, 1977).

Example 3: Complete cloning of the SNDH gene

(1) Construction of gene library of *G. oxydans* DSM 4025

The chromosomal DNA of *G. oxydans* DSM 4025 was prepared from the cells grown on M agar medium; 5% D-mannitol, 1.75% corn steep liquor, 5% baker's yeast, 0.25% $MgSO_4 \cdot 7H_2O$, 0.5% $CaCO_3$ (Pr.G.), 0.5% urea, and 2.0% agar (pH 7.0), for 4 days at 27°C. The chromosomal DNA (4 µg) was partially digested with 4 units of *Eco*R I in 20 µl of reaction mixture. A portion (8 µl) of the sample containing partially-digested DNA fragments was separated by an electrophoresis using 1% agarose gel. Fragments ranging from 15 to 35 kb were cut out and chemically melted to recover the fragments using QIAEX II (QIAGEN Inc., 28159 Avenue Stanford, Valencia, CA 91355, USA). The objective DNA fragments recovered were suspended in H_2O. On the other hand, 2 µg of a cosmid vector pVK100 was completely digested with *Eco*R I and dephosphorylated of the 5'-ends by treating of bacterial alkaline phosphatase (*E.coli* C75) (Takara Shuzo). The treated pVK100 (220 ng) was ligated with the15 - 35 kb *Eco*R I fragments (1 µg) using a ligation kit (Takara Shuzo) in 36 µl of reaction mixture. The ligated DNA which had been ethanol precipitated and resolved in appropriate volume of TE buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA) was used for *in vitro* packaging (Gigapack III Gold Packaging Extract, Stratagene, 11011 North Torrey Pines Road, La Jolla, CA 92037, USA) to infect *E. coli* VCS257, a host strain for the genomic library. As a result, totally 400,000 - 670,000 clones containing about 25 kb-inserted DNA fragments were obtained.

(2) Complete cloning of the SNDH gene by colony hybridization.

The probe that would be used for screening of the cosmid library described above to fish up clones carrying the complete SNDH gene by colony hybridization method, was constructed. The 41 bp DNA fragment encoding the N-terminal amino acid sequence of

SNDH was amplified and labeled by PCR-DIG labeling method (Roche Molecular Systems Inc., 1145 Atlantic Avenue, Alabama, CA94501, USA). PCR with plasmid pMTSN2 DNA as a template and oligonucleotide DNA primers, P13 (SEQ ID NO:8) and P14 (SEQ ID NO:9), was performed with thermostable taq polymerase (TAKARA Ex Taq™, Takara Shuzo Co., Ltd.), using a thermal cycler (Gene Amp PCR System 2400-R, PE Biosystems). The reaction was carried out with 25 cycles of 1) denaturation step at 94°C for 30 sec; 2) annealing step at 55°C for 30 sec; 3) synthesis step at 70°C for 1 min. Using the DIG-labeled probe, screening of the cosmid library (about 1,000 clones) by colony hybridization and chemiluminescent detection according to the method provided from the supplier (Roche Molecular Systems Inc., USA) was carried out. Consequently, three positive clones were isolated and one of them was designated pVSN5, which carried about 25 kb insert DNA in pVK100 vector. The DNA fragments of 3.2 kb *Eco*R I, 7.2 kb *Eco*R I, and *Pst* I 8.0 kb, which contain the upstream, the downstream, and the intact of the SNDH gene, respectively, were subcloned into pUC18 vector to obtain pUCSN19, pUCSN5 and pUCSNP4, respectively (Figure 2).

(3) Nucleotide sequencing of the SNDH gene

Plasmids pUCSN19, pUCSN5, and pUCSNP4 were used for nucleotide sequencing the region including the SNDH gene. Determined nucleotide sequence (SEQ ID NO:1; 3,408 bp) revealed that ORF of SNDH gene (1,827 bp, nucleotide positions at 258-2084 in SEQ ID NO:1) encoded the polypeptide of 609 amino acid residues (SEQ ID NO:2). Additional ORF, ORF-A, was found in the downstream of the SNDH ORF as illustrated in Figure 1. The ORF of ORF-A (1,101 bp, nucleotide positions at 2214-3314 of SEQ ID NO:1) encoded the polypeptide of 367 amino acids.

In the ORF of the SNDH gene, a signal peptide-like sequence (SEQ ID NO:4; 31 amino acids) is possibly included in the deduced amino acid sequence; it contains (i) many hydrophobic residues, (ii) positively-charged residues close to the N-terminus and (iii) Ala-Xaa-Ala motif as a cleaved signal sequence. The putative ribosome-binding site (Shine-Dalgarno, SD, sequence) for the SNDH gene was located at 6 bp upstream of the initiation codon (AGGAGA at nucleotide positions at 247-252 of SEQ ID NO:1).

Homology search for genes of the SNDH was performed with the program of FASTA in GCG (Genetics Computer Group, Madison, WI, USA). The 80% N-terminal region of the deduced amino acid sequence of the SNDH gene has moderate similarity with *A. calcoaceticus* soluble quinoprotein glucose dehydrogenase (GDH-B) at the identity of 41%. Further analysis by homology search in multiple alignment with the program of ClustalW revealed that two of very conserved regions around the presumed active site of *A. calco-*

aceticus GDH-B were found among hypothetical proteins defined by genome DNA analysis with the SNDH protein. The optional parameters for the analysis are as follows; MATRIX: blosum, GAPOPEN: 10.0, GAPEXT: 0.05, GAPDIST: 8, and any other parameter was set at default of the program. The result said the either amino acid sequences of positions at 224-231 and positions at 259-265 of SEQ ID NO:5 (the amino acid sequences of positions at 224-231 and/or 259-265 of SEQ ID NO:5 corresponds to the consensus sequences around the active site) were over 85% identical at the corresponding regions in the deduced amino acid sequences of AB013367 (*Bacillus halodurans* unknown protein), AE003996 (*Xylella fastidiosa* hypothetical protein), AE007222 (*Sinorhizobium meliloti* plasmid pSymA hypothetical protein), AE009889 (*Pyrobaculum aerophilum* hypothetical protein), AE004541 (*Pseudomonas aeruginosa* hypothetical protein), ECAE186 (*E.coli* K-12 hypothetical protein), AF472590 (*Sinorhizobium meliloti* hypothetical protein), P73001 (*Synechocystis sp.* hypothetical protein), and two real proteins of *A. calcoaceticus* GDH-B and the SNDH of this invention. Additionally it was found that the several conserved residues in the presumed active site in *A. calcoaceticus* GDH-B reported by Oubrie et al. [J. Mol. Biol. 289:319-333 (1999)], which were Arg227, Asn228, Gln230, Gly246, and Asp251 of SEQ ID NO:5, were completely conserved among those sequences. On the other hand, the rest 20% C-terminal region of the SNDH has moderate similarity with heme containing proteins such as c-type cytochrome, cytochrome f, from cyanobacterium *P. boryanum* and cd-1 type nitrite reductase, *nir* S, from a *Paracoccus denitrificans* strain. In the similar region (about 32% identity), a motif (Cys-Xaa-Xaa-Cys-His) defined as heme c binding was found at positions at 530-534 of SEQ ID NO:2. As a result shown in the above, the SNDH protein is thought to be one of quinohemoproteins from a genetically analysis.

Example 4: Expression of the SNDH gene in *E. coli*

Plasmids pUCSNP4 and pUCSNP9 (Figure 3), which have 8.0 kb *Pst* I-fragment containing the intact SNDH gene, were transformed into *E. coli* JM109 to confirm the expression and the activity of the SNDH proteins.

The conversion activity of L-sorbosone to vitamin C by using cytosol fraction of the recombinant *E. coli* was tested (Table 1). The cytosol fraction was prepared by ultracentrifugation (100,000 x *g*, 45 min) of the cell free extract in 50 mM potassium phosphate buffer (pH 7.0). The reaction mixture (100 µl) consisted of 125 µg of cytosol fraction of the recombinant *E. coli*, 50 mM of L-sorbosone, 1.0 mM of phenazine mesosulfate (PMS), and additionally 1.0 µM of PQQ and 1.0 mM of $CaCl_2$ as cofactors depending on a case. The enzyme reaction was carried out at 30°C for 30 minutes. Native holo-SNDHs of the cells

cultivated in LB medium containing 10 µM of PQQ and 1.0 mM of $CaCl_2$ produced vitamin C definitely under the defined reaction condition without the cofactors of PQQ and $CaCl_2$. In addition of the cofactors, the apo-enzyme from pUCSNP4 and pUCSNP9 showed almost the same activity as those of the native holo-enzyme.

Table 1:

microorganism	PQQ and $CaCl_2$ in the medium	Specific activity (mU/mg Protein)	
		+ PQQ and $CaCl_2$	- PQQ and $CaCl_2$
E. coli JM109/pUCSNP4	+	0.187	0.224
E.coli JM109/pUCSNP9	+	0.198	0.252
E. coli JM109/pUC18	+	0.000	0.000
E. coli JM109/pUCSNP4	-	0.155	0.000
E. coli JM109/pUCSNP9	-	0.176	0.000
E. coli JM109/pUC18	-	0.000	0.000
G. oxydans DSM 4025	-	0.026	0.026

One unit (U) of the enzyme was defined as the amount of enzyme, which produces 1.0 mg of vitamin C in the defined reaction.

Example 5: Construction and cultivation of SNDH-gene disruptants of *G. oxydans* strains

Figure 4 shows the scheme for the construction of SNDH gene targeting vector, GOMTR1SN::Km (SNDH-disruptant). First, plasmid pSUPSN was constructed by a ligation of 8.0 kb *Pst* I fragment containing the SNDH gene from plasmid pUCSNP4 with a suicide vector pSUP202. Second, a kanamycin-resistant-gene cassette (Km cassette) was inserted into the *Eco*R I site of the SNDH gene cloned in plasmid pSUPSN to obtain plasmid pSUPSN::Km (Km^rTc^r). Then, plasmid pSUPSN::Km was introduced into GOMTR1, which was a rifampicin (Rif) resistant derived spontaneously from wild *G. oxydans* DSM 4025 strain, to obtain SNDH-null mutants ($Km^rRif^rTc^s$).

G. oxydans GOMTR1 was cultivated in a 200 ml flask containing 50 ml of T broth, which was composed of 30 g/l of Trypticase Soy Broth (BBL; Becton Dickinson and Company, Cockeysville, MD 21030, USA) and 3 g/l of yeast extract (Difco; Becton Dickinson Microbiology Systems, Becton Dickinson and Company, Sparks, MD 21152, USA) with 100 µg/ml of rifampicin at 30°C overnight. *E. coli* HB101 (pRK2013) [D. H. Figurski, Proc. Natl. Acad. Sci. USA, 76, 1648-1652, 1979] and *E. coli* JM109 (pSUPSN::Km) were cultivated in test tubes containing 2 ml of LB medium with 50 µg/ml of kanamycin at 30°C

overnight. Cultured cells of GOMTR1, *E. coli* HB101 (pRK2013), and *E. coli* JM109 (pSUPSN::Km) were collected separately by centrifugation and suspended in LB medium at the OD of about 20, 2, and 2, respectively. Then these cell suspensions were mixed at the same volume and the mixture was spread out on a 0.45 μm nitrocellulose membrane (PROTRAN, Schleicher & Schuell GmbH, Postfach 4, D-37582 Dassel, Germany) put on an agar medium, which were composed of 5.0% mannitol, 0.25% $MgSO_4 \cdot 7H_2O$, 1.75% corn steep liquor, 5.0% baker's yeast, 0.5% urea, 0.5% $CaCO_3$, and 2.0% agar, to do conjugal transfer the suicide plasmid from the *E. coli* donor to GOMTR1. After cultivation at 27°C for a day, the cells containing transconjugants were suspended and diluted appropriately with T broth, and spread out on the screening agar plates containing 100 μg/ml of rifampicin and 50 μg/ml of kanamycin. Finally, several objective transconjugants ($Km^r Rif^r Tc^s$) which had the disrupted SNDH gene with Km cassette were obtained.

GOMTR1 and the disruptants, GOMTR1SN::Km, were grown on an agar plate containing 8.0% L-sorbose, 0.25% $MgSO_4 \cdot 7H_2O$, 1.75% corn steep liquor, 5.0% baker's yeast, 0.5% urea, 0.5% $CaCO_3$, and 2.0% agar at 27°C for 4 days. One loopful of the cells was inoculated into 50 ml of a seed culture medium (pH 6.0) containing 4% D-sorbitol, 0.4% yeast extract, 0.05% glycerol, 0.25% $MgSO_4 \cdot 7H_2O$, 1.75% corn steep liquor, 0.1% urea, and 1.5% $CaCO_3$ in a 500 ml Erlenmeyer flask, and cultivated at 30°C with 180 rpm for one day on a rotary shaker. The seed culture thus prepared was used for inoculating 50 ml of a main culture medium, which composed of 12.0% L-sorbose, 2.0% urea, 0.05% glycerol, 0.25% $MgSO_4 \cdot 7H_2O$, 3.0% corn steep liquor, 0.4% yeast extract, and 1.5% $CaCO_3$ in a 500 ml Erlenmeyer flask. The cultivation was carried out at 30°C and 180 rpm for 4 days. As shown in Table 2, the SNDH-gene disruptants had about 3% higher of the molar conversion yield against the substrate consumed than that of the parent strain.

Table 2

Strain	2KGA (g/L)	Residual L-sorbose (g/L)	*Molar yield (mol %)
GOMTR1SN::Km	96.7	15.3	99.2
GOMTR1	98.8	9.8	95.5

*Molar yield: mol 2-KGA produced/mol L-sorbose consumed.

Example 6: Introduction of the plasmids carrying the SNDH gene into the SNDH-gene disruptant of *G. oxydans* DSM 4025

Several kinds of SNDH-expression plasmids using broad host range vector pVK100 were constructed as shown in Figure 5. Those plasmids have different insert DNAs at the *Hind* III site of pVK100 described as follows: pVSN117 has the insert DNA containing the incomplete SNDH gene encoding a polypeptide until Gly535 of SEQ ID NO:5, i.e. a C-terminal deleted SNDH gene, which expresses only 55 kDa protein, pVSN106 and pVSN114 have the insert DNA containing the intact SNDH gene. Those plasmids were introduced into strain GOMTR1SN::Km, SNDH-gene disruptant derived from *G. oxydans* DSM 4025, by conjugal transfer method.

These transconjugants having the plasmids shown in Figure 5 were grown on an agar plate containing 10.0% L-sorbose, 0.25% $MgSO_4 \cdot 7H_2O$, 1.75% corn steep liquor, 5.0% baker's yeast, 0.5% urea, 0.5% $CaCO_3$, and 2.0% agar at 27°C for 4 days. The enzyme reaction mixture consisted of 80 µg of cell free extract of the recombinant *Gluconobacter* strains, 25 mM potassium phosphate buffer (pH 7.0), 50 mM of L-sorbosone, and 0.05 mM of PMS. The enzyme reaction was carried out at 30°C for 30 min with shaking at 1,000 rpm. The result is shown in Table 3.

Table 3

Host cell	Vector DNA	Vitamin C produced (mg/L)
GOMTR1SN::Km-2	pVK100	0.0
	pVSN117	473.2
	pVSN106	845.3
	pVSN114	860.2

Claims

1. An isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide sequence at least 95% identical to the nucleotide sequence of SEQ ID NO:1.

2. An isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide sequence at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO:1, (b) nucleotides 351-2084 of SEQ ID NO:1, (c) nucleotides 258-1955 of SEQ ID NO:1, and (d) nucleotides 351-1955 of SEQ ID NO:1.

3. An isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide consisting of amino acids of SEQ ID NO:2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609 of SEQ ID NO:2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO:2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO:2.

4. An isolated nucleic acid molecule encoding a polypeptide having aldehyde dehydrogenase activity, wherein the complement of said nucleic acid molecule hybridizes under standard conditions with the nucleic acid molecule of any one of claims 1 to 3.

5. An expression vector which comprises the nucleic acid molecule of any one of claims 1 to 4.

6. The expression vector of claim 5, wherein said vector is derived from pQE-plasmids, pUC-plasmids, pBluescript II, pACYC177, pACYC184, and their derivative plasmids, and a broad host range plasmid such as pVK100 and RSF1010.

7. A recombinant microorganism which is transformed with the expression vector of claim 5 or 6.

8. The recombinant microorganism of claim 7, wherein said microorganism comprises the nucleic acid molecule of any one of claims 1 to 4 on its chromosomal DNA.

9. The recombinant microorganism of claim 7 or 8, wherein said microorganism is selected from the group consisting of bacterial cells, yeast cells, and plant cells.

10. The recombinant microorganism of claim 9, wherein said microorganism is a member of the genus selected from the group consisting of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Klebsiella*, *Acinetobacter*, and *Escherichia*.

11. A process for the production of vitamin C and/or 2-KGA from L-sorbosone comprising (a) propagating the recombinant microorganism of claim 7 in an appropriate culture media, and (b) recovering and separating vitamin C and/or 2-KGA from said culture media.

12. A process for the production of vitamin C and/or 2-KGA from L-sorbosone comprising (a) propagating a recombinant organism in an appropriate culture media, wherein the nucleic acid molecule of any one of claims 1 to 4 is heterologously introduced to said recombinant organism, and (b) recovering and separating vitamin C and/or 2-KGA from said culture media.

13. A process for the production of vitamin C and/or 2-KGA from L-sorbosone comprising (a) propagating a recombinant organism in an appropriate culture media, wherein a nucleic acid molecule comprising a polynucleotide encoding a polypeptide whose consensus amino acid sequences around the active site are at least 85% identical to those of the polypeptide of SEQ ID NO:5 is heterologously introduced to said recombinant organism, and (b) recovering and separating vitamin C and/or 2-KGA from said culture media.

14. A process for the production of 2-KGA via L-sorbosone from an appropriate sugar compound comprising (a) propagating a microorganism belonging to *Gluconobacter oxydans* DSM 4025 in an appropriate culture media, wherein a gene encoding aldehyde dehydrogenase encoded by any one of claims 1 to 4 is disrupted in said microorganism, and (b) recovering and separating 2-KGA from said culture media.

15. A process according to claim 14, wherein the sugar compound is selected from the group consisting of L-sorbosone, D-glucose, D-sorbitol, and L-sorbose.

16. A process for the production of aldehyde dehydrogenase encoded by any one of claims 1 to 4, the process comprising (a) propagating the recombinant microorganism of claim 7 in an appropriate culture media, and (b) recovering and separating said aldehyde dehydrogenase from said culture media.

17. A process for the production of aldehyde dehydrogenase encoded by any one of claims 1 to 4, the process comprising (a) propagating a recombinant organism in an appropriate

culture media, wherein the nucleic acid molecule of any one of claims 1 to 4 is heterologously introduced to said recombinant organism, and (b) recovering and separating said aldehyde dehydrogenase from said culture media.

18. A process for the production of aldehyde dehydrogenase encoded by any one of claims 1 to 4, the process comprising (a) propagating a recombinant organism in an appropriate culture media, wherein a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide whose consensus amino acid sequences around the active site are at least 85% identical to those of the polypeptide of SEQ ID NO:5 is heterologously introduced to said recombinant organism, and (b) recovering and separating said aldehyde dehydrogenase from said culture media.

***

SEQUENCE LISTING

<110> Roche Vitamins AG

<120> A gene encoding aldehyde dehydrogenase and use thereof

<130> NDR5237

<140>

<141>

<160> 9

<170> PatentIn Ver. 2.1

<210> 1

<211> 3408

<212> DNA

<213> Gluconobacter oxydans

<220>

<221> CDS

<222> (258)..(2087)

<220>

<221> CDS

<222> (2214)..(3317)

<400> 1

GCGACTGGCA GCAGCGCAAC TATGACCACT ATGGCCTGCC GCCCTATTGG 50

ATCTAACTGA TCCAGTAAGC CACCATCAGC CGGCCCCTGC GGGGGCCGGC 100

TTTTTGCGCT AGACCCCGCC GAGGTGCTGT CGTAACCTAA GGTCACATCT 150

TTACTTCCAC ATCCGCCCTT GTCAGTTCTG ACGTGACAAA TTGTCGCGGT 200

CATGCTGCTG AATGCGGATG CCAGTCCCAG ATCCAAGCCC GACGCAAGGA 250

GACGTAGATG TTACCCAAAT CATTGAAACA TAAGAATGGC GCCATGCGCC 300

TTGTCGCAGC CTCGACCCTT GCGCTGATGA TCGGCGCGGG TGCCCATGCG 350

CAGGTAAACC CGGTCGAAGT GCCGGTGGGC GCGAACGAGA CCTTTACCTC 400

GCGCGTGCTG ACCACCGGCC TGTCGAACCC TTGGGAAATC ACCTGGGGCC 450

CCGACAATAT GCTGTGGGTG ACCGAGCGAT CTTCCGGCGA AGTGACGCGC 500

GTCGACCCCA ATACCGGCGA GCAGCAGGTC CTGCTGACCC TGACCGATTT 550

CAGCGTCGAT GTGCAACACC AGGGCCTACT TGGCCTCGCG CTGCATCCTG 600

AGTTTATGCA AGAGAGCGGC AACGACTACG TCTATATCGT CTACACTTAT 650

AACACCGGCA CCGAAGAAGC GCCCGATCCG CATCAAAAGC TGGTGCGTTA 700

TGCCTATGAC GCTGCCGCGC AGCAGCTGGT CGATCCGGTT GATCTGGTCG 750

CAGGCATTCC CGCAGGCAAC GACCACAATG GCGGTCGCAT CAAATTCGCC 800

CCCGATGGCC AACACATCTT TTACACGCTG GGCGAGCAAG GCGCGAACTT 850

TGGCGGTAAC TTCCGCCGTC CGAACCACGC GCAACTGCTG CCGACGCAAG 900

AGCAGGTCGA CGCGGGCGAT TGGGTCGCCT ATTCGGGCAA GATCCTGCGC 950

GTGAACCTTG ACGGCACGAT CCCCGAAGAC AACCCCGAGA TCGAGGGCGT 1000

GCGTAGCCAT ATCTTTACCT ATGGCCACCG TAACCCGCAG GGCATCACCT 1050

TTGGCCCCGA CGGCACCATT TATGCCACCG AACACGGCCC CGATACGGAT 1100

GACGAGCTGA ACATCATCGC CGGCGGTGGC AACTATGGGT GGCCGAATGT 1150

GGCCGGCTAT CGCGATGGCA AATCCTATGT CTACGCTGAT TGGAGCCAAG 1200
CGCCCGCTGA CCAGCGTTAC ACCGGTCGCG CCGGTATCCC CGACACCGTG 1250
CCGCAATTCC CCGAGCTGGA ATTCGCGCCC GAGATGGTCG ATCCGCTGAC 1300
AACCTATTGG ACGGTGGATA ATGATTACGA TTTCACCGCC AATTGCGGCT 1350
GGATCTGTAA TCCGACGATC GCGCCTTCGT CTGCCTATTA CTATGCGGCG 1400
GGCGAGAGCG GTATCGCGGC TTGGGATAAT TCGATCCTGA TCCCGACGCT 1450
GAAACATGGC GGCATCTATG TGCAGCACCT CAGCGATGAT GGCCAATCTG 1500
TCGACGGCCT GCCCGAGCTG TGGTTCAGCA CCCAGAACCG CTATCGCGAT 1550
ATCGAGATCA GCCCCGATAA CCATGTTTTT GTGGCGACCG ACAACTTTGG 1600
CACCTCGGCG CAGAAATATG GCGAGACCGG CTTTACCAAC GTGCTGCATA 1650
ACCCCGGCGC GATCCTTGTC TTTAGCTATG TCGGCGAGGA TGCTGCGGGT 1700
CAGACCGGAA TGATGACCGC GCCCGCACCG CAGACGCAAT ACACGCAAGT 1750
GCCCGCCGAG GGTGCAGGCG CGGGCGCGAC TGAGGTTGCG GATGTCGATT 1800
ACGACACGCT GTTCACCGAA GGCCAGACCC TTTATGGCAG CGCATGTGCC 1850
GCGTGCCATG GTGCCGCTGG CCAAGGTGCG CAGGGCCCGA CCTTTGTGGG 1900
CGTGCCGGAT GTGACGGGTG ACAAGGACTA CCTTGCCCGC ACCATCATCC 1950
ACGGTTTTGG CTATATGCCG TCGTTTGCGA CTCGGCTGGA TGACGAGGAG 2000
GTTGCCGCCA TCGCGACCTT TATCCGCAAC AGCTGGGGCA ATGACGAAGG 2050
CATCCTGACC CCGGCCGAGG CCGCTGCCAC CCGCTGAATG CTGTAAAAAC 2100
CACCCTCGCC TGCACATCAG GCGGGGGTAT TTCATTTATT TTCACATCTG 2150
CCTTTGACAT GTGCCGCTAT CACGGTTAAT GCGGCCCTTC GGCTGTTCTG 2200
GGTCTAAGCG GGTGTGTTGC CCGATAAGAG AGACGGTTCA GTCCCTCCCG 2250
CCCTATTTAG GGCCCATTTA GGCAGAATAG TTTTGACTCA TCAAAATATC 2300
GCCGCGCCTC TGGCCGCGGC CCTTTCGCAA CGTGGATATG AAACGCTGAC 2350
CGCCGTGCAG CAAGCTGTGC TTGCGCCCGA GGCTGATGGC CGCGACCTGC 2400
TGGTGTCGGC ACAGACCGGT TCGGGTAAGA CGGTGGCCTT TGGTATCGCA 2450
GTCGCGCCCG ACCTTTTGGG CGACGACAAT ATCCTGCCGC TGAACACGCC 2500
GCCTGTTGCG CTGTTCATCG CCCCCACGCG CGAGCTTGCG CTGCAAGTTG 2550
CTCAGGAACT GACCTGGCTT TACGCCAATG CAGGTGCCCA GATCGCGACC 2600
TGCGTCGGCG GTATGGATTA CCGCACCGAG CGCCGCGCCC TTGCACGTCT 2650
GCCGCAAATC GTTGTCGGCA CGCCCGGCCG TCTGCGCGAC CATATCGACC 2700
GTGGCGGCCT TGACCTGTCC GAATTGCGCG TGACCGTGCT GGACGAAGCG 2750
GATGAGATGC TCGACCTCGG CTTCCGCGAT GATCTGCAAT ATATCTTGCA 2800
AGCCGCGCCC GAAGATCGCC GCACGCTGAT GTTCTCGGCC ACCGTGCCGC 2850
GCGAGATTGA AAAACTGGCC CGCGACTTCC AAAATGACGC CCTGCGTCTG 2900
GAAACCCGTG GCGAGGCCAA GCAGCACAAC GACATCAGCT ACCAAGCTTT 2950
GTCGGTCACC ATGCGCGATC GCGAAAACGC CATTTTCAAC ATGCTGCGTT 3000
TTTATGAATC GCGCACGGCG ATCATCTTCT GCAAGACCCG CGCCAATGTG 3050
AATGATCTGC TGTCGCGGAT GAGCGGTCGT GGCTTCCGCG TGGTGGCCCT 3100
GTCGGGCGAG CTGTCGCAAC AGGAACGCAC CAACGCGCTG CAAGCGCTGC 3150
GTGATGGCCG CGCCAACGTT TGTATCGCGA CCGACGTCGC GGCGCGCGGC 3200
ATTGACTTGC CGGGCCTCGA GCTGGTGATC CACTACGATC TGCCGACCAA 3250
TGCCGAAACC CTGCTGCACC GCTCGGGCCG TACCGGCCGC CGGGTGCCAA 3300
GGGCGTCTCG GCGCTGATCG TCACCCCCGG CGATTTCAAA AAAGCGCAGC 3350

GTTTGCTGAG CTTTGCCAAA GTGACCGCGG AATGGGGCAA GGCGCCTTCG 3400
GCCGAAGA 3408

<210> 2
<211> 609
<212> PRT
<213> Gluconobacter oxydans
<220>
<221> SIGNAL
<222> (1)..(31)
<220>
<221> CHAIN
<222> (32)..(609)

<400> 2

Met Leu Pro Lys Ser Leu Lys His Lys Asn Gly Ala Met Arg Leu 15

Val Ala Ala Ser Thr Leu Ala Leu Met Ile Gly Ala Gly Ala His 30

Ala Gln Val Asn Pro Val Glu Val Pro Val Gly Ala Asn Glu Thr 45

Phe Thr Ser Arg Val Leu Thr Thr Gly Leu Ser Asn Pro Trp Glu 60

Ile Thr Trp Gly Pro Asp Asn Met Leu Trp Val Thr Glu Arg Ser 75

Ser Gly Glu Val Thr Arg Val Asp Pro Asn Thr Gly Glu Gln Gln 90

Val Leu Leu Thr Leu Thr Asp Phe Ser Val Asp Val Gln His Gln 105

Gly Leu Leu Gly Leu Ala Leu His Pro Glu Phe Met Gln Glu Ser 120

Gly Asn Asp Tyr Val Tyr Ile Val Tyr Thr Tyr Asn Thr Gly Thr 135

Glu Glu Ala Pro Asp Pro His Gln Lys Leu Val Arg Tyr Ala Tyr 150

Asp Ala Ala Ala Gln Gln Leu Val Asp Pro Val Asp Leu Val Ala 165

Gly Ile Pro Ala Gly Asn Asp His Asn Gly Gly Arg Ile Lys Phe 180

Ala Pro Asp Gly Gln His Ile Phe Tyr Thr Leu Gly Glu Gln Gly 195

Ala Asn Phe Gly Gly Asn Phe Arg Arg Pro Asn His Ala Gln Leu 210

Leu Pro Thr Gln Glu Gln Val Asp Ala Gly Asp Trp Val Ala Tyr 225

Ser Gly Lys Ile Leu Arg Val Asn Leu Asp Gly Thr Ile Pro Glu 240

Asp Asn Pro Glu Ile Glu Gly Val Arg Ser His Ile Phe Thr Tyr 255

Gly His Arg Asn Pro Gln Gly Ile Thr Phe Gly Pro Asp Gly Thr 270

Ile Tyr Ala Thr Glu His Gly Pro Asp Thr Asp Asp Glu Leu Asn 285

Ile Ile Ala Gly Gly Gly Asn Tyr Gly Trp Pro Asn Val Ala Gly 300

Tyr Arg Asp Gly Lys Ser Tyr Val Tyr Ala Asp Trp Ser Gln Ala 315

Pro Ala Asp Gln Arg Tyr Thr Gly Arg Ala Gly Ile Pro Asp Thr 330

Val Pro Gln Phe Pro Glu Leu Glu Phe Ala Pro Glu Met Val Asp 345

Pro Leu Thr Thr Tyr Trp Thr Val Asp Asn Asp Tyr Asp Phe Thr 360

Ala Asn Cys Gly Trp Ile Cys Asn Pro Thr Ile Ala Pro Ser Ser 375

Ala Tyr Tyr Tyr Ala Ala Gly Glu Ser Gly Ile Ala Ala Trp Asp 390

Asn Ser Ile Leu Ile Pro Thr Leu Lys His Gly Gly Ile Tyr Val 405

Gln His Leu Ser Asp Asp Gly Gln Ser Val Asp Gly Leu Pro Glu 420

Leu Trp Phe Ser Thr Gln Asn Arg Tyr Arg Asp Ile Glu Ile Ser 435

Pro Asp Asn His Val Phe Val Ala Thr Asp Asn Phe Gly Thr Ser 450

Ala Gln Lys Tyr Gly Glu Thr Gly Phe Thr Asn Val Leu His Asn 465

Pro Gly Ala Ile Leu Val Phe Ser Tyr Val Gly Glu Asp Ala Ala 480

Gly Gln Thr Gly Met Met Thr Ala Pro Ala Pro Gln Thr Gln Tyr 495

Thr Gln Val Pro Ala Glu Gly Ala Gly Ala Gly Ala Thr Glu Val 510

Ala Asp Val Asp Tyr Asp Thr Leu Phe Thr Glu Gly Gln Thr Leu 525

Tyr Gly Ser Ala Cys Ala Ala Cys His Gly Ala Ala Gly Gln Gly 540

Ala Gln Gly Pro Thr Phe Val Gly Val Pro Asp Val Thr Gly Asp 555

Lys Asp Tyr Leu Ala Arg Thr Ile Ile His Gly Phe Gly Tyr Met 570

Pro Ser Phe Ala Thr Arg Leu Asp Asp Glu Glu Val Ala Ala Ile 585

Ala Thr Phe Ile Arg Asn Ser Trp Gly Asn Asp Glu Gly Ile Leu 600

Thr Pro Ala Glu Ala Ala Ala Thr Arg 609

<210> 3
<211> 14
<212> PRT
<213> Gluconobacter oxydans

<400> 3
Gln [Xaa/Gly] Asn [Pro/Lys] Val Glu Val Pro Val Gly Ala Asn Glu Thr 14

<210> 4
<211> 31
<212> PRT
<213> Gluconobacter oxydans

<220>
<221> SIGNAL
<222> (1)..(31)

<400> 4
Met Leu Pro Lys Ser Leu Lys His Lys Asn Gly Ala Met Arg Leu 15

Val Ala Ala Ser Thr Leu Ala Leu Met Ile Gly Ala Gly Ala His 30

Ala 31

<210> 5
<211> 578
<212> PRT
<213> Gluconobacter oxydans

<220>
<221> CHAIN

<222> (1)..(578)

<400> 5

Gln Val Asn Pro Val Glu Val Pro Val Gly Ala Asn Glu Thr Phe 15

Thr Ser Arg Val Leu Thr Thr Gly Leu Ser Asn Pro Trp Glu Ile 30

Thr Trp Gly Pro Asp Asn Met Leu Trp Val Thr Glu Arg Ser Ser 45

Gly Glu Val Thr Arg Val Asp Pro Asn Thr Gly Glu Gln Gln Val 60

Leu Leu Thr Leu Thr Asp Phe Ser Val Asp Val Gln His Gln Gly 75

Leu Leu Gly Leu Ala Leu His Pro Glu Phe Met Gln Glu Ser Gly 90

Asn Asp Tyr Val Tyr Ile Val Tyr Thr Tyr Asn Thr Gly Thr Glu 105

Glu Ala Pro Asp Pro His Gln Lys Leu Val Arg Tyr Ala Tyr Asp 120

Ala Ala Ala Gln Gln Leu Val Asp Pro Val Asp Leu Val Ala Gly 135

Ile Pro Ala Gly Asn Asp His Asn Gly Gly Arg Ile Lys Phe Ala 150

Pro Asp Gly Gln His Ile Phe Tyr Thr Leu Gly Glu Gln Gly Ala 165

Asn Phe Gly Gly Asn Phe Arg Arg Pro Asn His Ala Gln Leu Leu 180

Pro Thr Gln Glu Gln Val Asp Ala Gly Asp Trp Val Ala Tyr Ser 195

Gly Lys Ile Leu Arg Val Asn Leu Asp Gly Thr Ile Pro Glu Asp 210

Asn Pro Glu Ile Glu Gly Val Arg Ser His Ile Phe Thr Tyr Gly 225

His Arg Asn Pro Gln Gly Ile Thr Phe Gly Pro Asp Gly Thr Ile 240

Tyr Ala Thr Glu His Gly Pro Asp Thr Asp Asp Glu Leu Asn Ile 255

Ile Ala Gly Gly Gly Asn Tyr Gly Trp Pro Asn Val Ala Gly Tyr 270

Arg Asp Gly Lys Ser Tyr Val Tyr Ala Asp Trp Ser Gln Ala Pro 285

Ala Asp Gln Arg Tyr Thr Gly Arg Ala Gly Ile Pro Asp Thr Val 300

Pro Gln Phe Pro Glu Leu Glu Phe Ala Pro Glu Met Val Asp Pro 315

Leu Thr Thr Tyr Trp Thr Val Asp Asn Asp Tyr Asp Phe Thr Ala 330

Asn Cys Gly Trp Ile Cys Asn Pro Thr Ile Ala Pro Ser Ser Ala 345

Tyr Tyr Tyr Ala Ala Gly Glu Ser Gly Ile Ala Ala Trp Asp Asn 360

Ser Ile Leu Ile Pro Thr Leu Lys His Gly Gly Ile Tyr Val Gln 375

His Leu Ser Asp Asp Gly Gln Ser Val Asp Gly Leu Pro Glu Leu 390

Trp Phe Ser Thr Gln Asn Arg Tyr Arg Asp Ile Glu Ile Ser Pro 405

Asp Asn His Val Phe Val Ala Thr Asp Asn Phe Gly Thr Ser Ala 420

Gln Lys Tyr Gly Glu Thr Gly Phe Thr Asn Val Leu His Asn Pro 435

Gly Ala Ile Leu Val Phe Ser Tyr Val Gly Glu Asp Ala Ala Gly 450

Gln Thr Gly Met Met Thr Ala Pro Ala Pro Gln Thr Gln Tyr Thr 465

Gln Val Pro Ala Glu Gly Ala Gly Ala Gly Ala Thr Glu Val Ala 480

Asp Val Asp Tyr Asp Thr Leu Phe Thr Glu Gly Gln Thr Leu Tyr 495

Gly Ser Ala Cys Ala Ala Cys His Gly Ala Ala Gly Gln Gly Ala 510

Gln Gly Pro Thr Phe Val Gly Val Pro Asp Val Thr Gly Asp Lys 525

Asp Tyr Leu Ala Arg Thr Ile Ile His Gly Phe Gly Tyr Met Pro 540

Ser Phe Ala Thr Arg Leu Asp Asp Glu Glu Val Ala Ala Ile Ala 555

Thr Phe Ile Arg Asn Ser Trp Gly Asn Asp Glu Gly Ile Leu Thr 570

Pro Ala Glu Ala Ala Ala Thr Arg 578

<210> 6
<211> 17
<212> DNA
<213> Artificial Sequence

<220>

<223> an artificially synthesized primer sequence

<400> 6
carggyaacc csgtbga 17

<210> 7
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> an artificially synthesized primer sequence

<220>
<221> misc_feature
<222> 9
<223> n is a or g or c or t

<400> 7
gtytcgttng crccvac 17

<210> 8
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> an artificially synthesized primer sequence

<400> 8
cagggtaacc cggtc 15

<210> 9
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> an artificially synthesized primer sequence

<400> 9
gactcgtttg cgccc 15

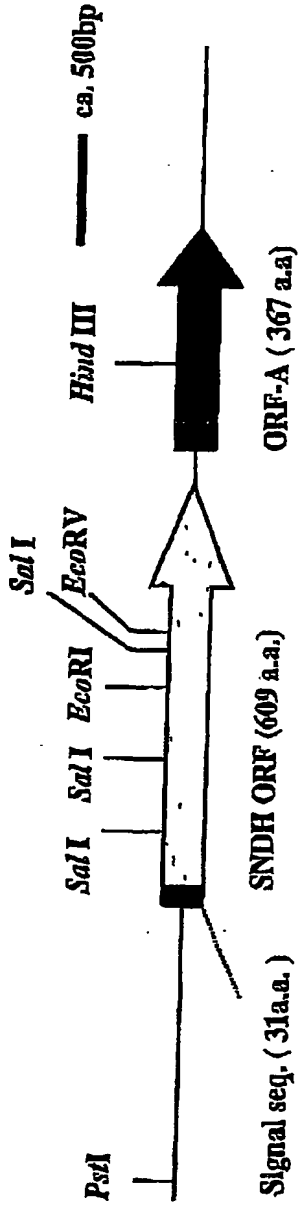

ca. 500bp
Hind III
ORF-A (367 a.a)
Sal I
EcoRV
EcoRI
Sal I
Sal I
SNDH ORF (609 a.a.)
PstI
Signal seq. (31a.a.)

Fig.1.

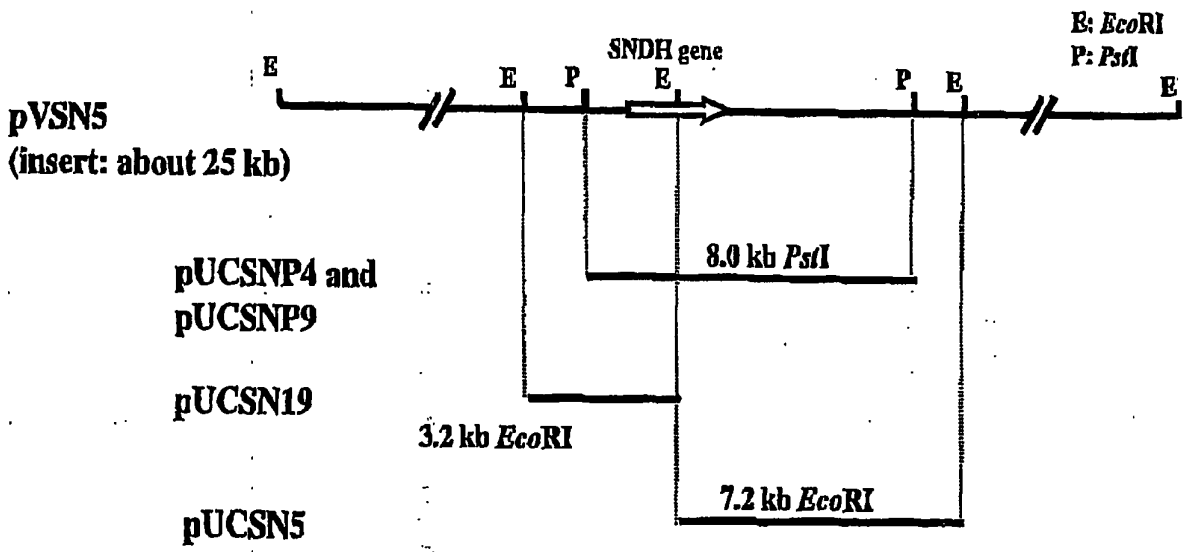
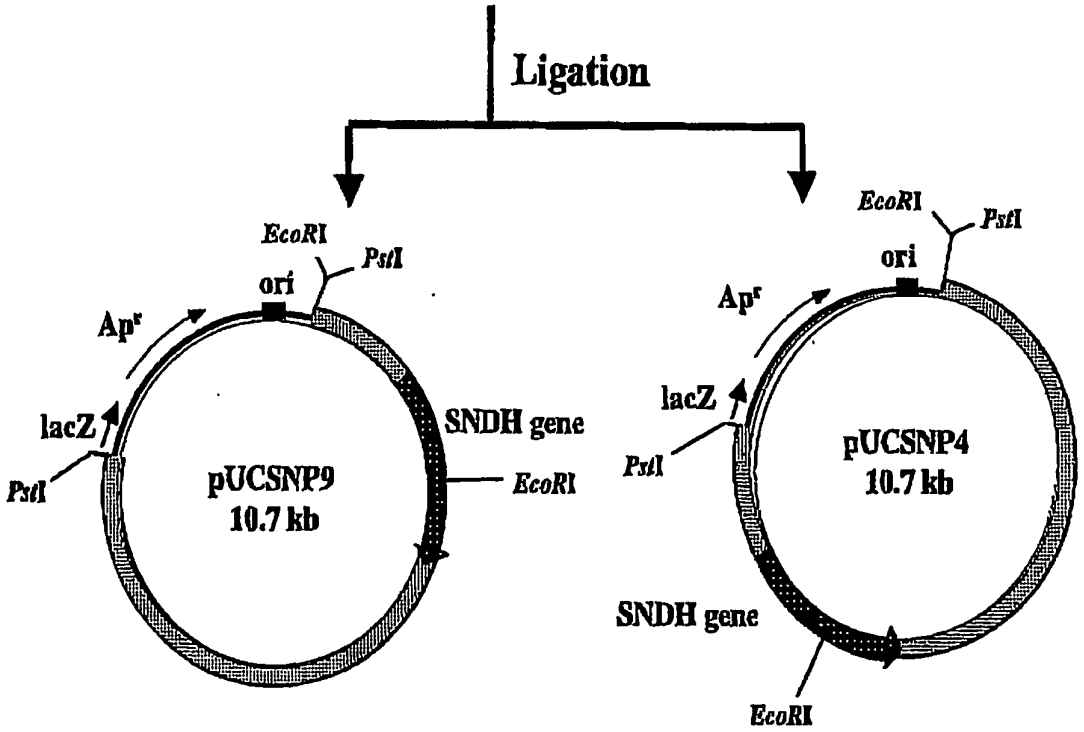

E: EcoRI
P: PstI
SNDH gene
E
E
P
E
P
E
E
pVSN5
(insert: about 25 kb)
8.0 kb PstI
pUCSNP4 and
pUCSNP9
pUCSN19
3.2 kb EcoRI
7.2 kb EcoRI
pUCSN5

Fig. 2.

- 8.0 kb *PstI* fragment including intact SNDH gene from pVSN5
- pUC18 plasmid vector digested with *PstI*



SNDH ORF-A NEW

Fig. 3.

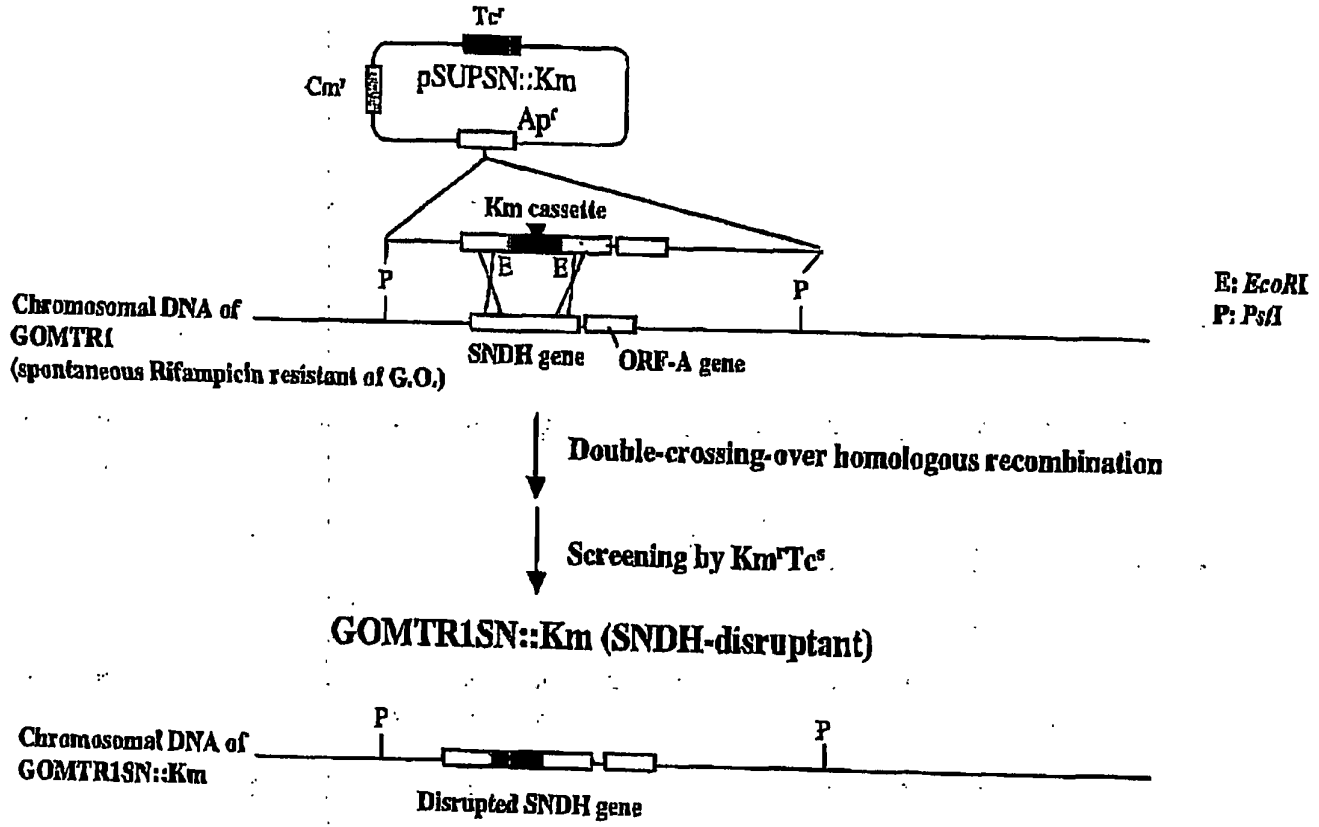

Tc^r
Cm^r
pSUPSN::Km
Ap^r
Km cassette
P
E
E
P
Chromosomal DNA of
GOMTR1
(spontaneous Rifampicin resistant of G.O.)
SNDH gene
ORF-A gene
E: EcoRI
P: PstI
Double-crossing-over homologous recombination
Screening by Km^rTc^s
GOMTR1SN::Km (SNDH-disruptant)
P
P
Chromosomal DNA of
GOMTR1SN::Km
Disrupted SNDH gene

FIG. 4.

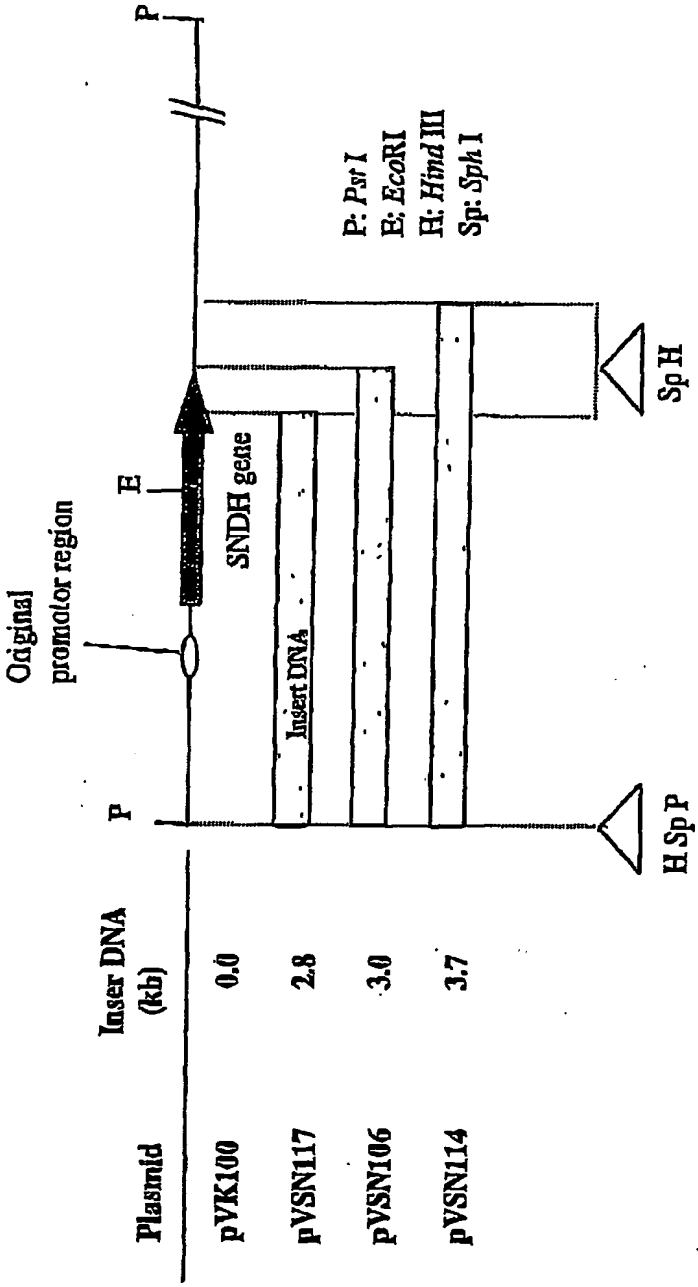

P
P: Psr I
E: EcoRI
H: Hind III
Sp: Sph I
Sp H
SNDH gene
E
Original promotor region
Insert DNA
P
H Sp P
Inser DNA (kb)
Plasmid
pVK100
0.0
pVSN117
2.8
pVSN106
3.0
pVSN114
3.7

Fig. 5.